Exploring Nebraska's Decision Rules for Adequate Yearly Progress

A Technical Report
Prepared
for

Dr. Douglas D. Christensen
Commissioner of Education

Nebraska Department of Education

May, 2004

EXECUTIVE SUMMARY

Findings

- A strong relationship existed between the number of AYP judgments and probability of failing to make AYP
- Minimum n-count changes beyond current levels had a minimal effect on AYP status change
- Confidence intervals significantly reduced misclassification across the participation rate indicator
- Confidence intervals reduced performance and other academic indicator misclassifications for elementary schools
- High and middle schools performance distances to the applicable thresholds combined with larger subgroup n-counts maintained their failed cell distributions
- Other academic indicators had a disproportional effect on AYP status
- Significant performance gaps existed between subgroups, which were not compensated by confidence intervals
- Students with disabilities performance gap may have a significantly negative impact on future AYP status

Future Considerations

- Consider improvement designs as alternatives to "Safe Harbor"
- Explore the school year 2004 performance of subgroups across Nebraska
- Conduct research on changes in AYP status across demographic and other factors
- Analyze the impact of additional grades and increased AYP thresholds
- Identify and modify decision-making rules with limited educational value (*i.e.,* requiring same subject, same subgroup criterion for school choice)

Research Context

The decision rules associated with the No Child Left Behind Act of 2001 (NCLB) were implemented for the first time in Nebraska. These decision rules are part of a conjunctive accountability system required under the federal statutes, which has been shown to misclassify schools and local educational agencies adequate yearly progress (AYP) status. Hill and DePascale (2003) have determined that sampling error has a major influence on the reliability statistic of accountability systems. Reliability has been defined by the aforementioned researchers as the probability of misclassifying a school (i.e. failing to make AYP). The misclassification of schools that results in corrective actions (labels, school choice, supplemental services, etc.) poses a credible validity threat to the system.

Efforts to reduce the standard error and subsequently the misclassification of schools may inadvertently generate a threat to the construct validity of the system. In other words, if the system is attempting to measure the academic productivity of subgroups but most subgroup sizes fail to reach the minimum n-count threshold, the accountability system will fail to operate in a coherent manner. An accountability system by which a school's size and diversity influences the number of judgments (and the probability of misclassifications) is operating incoherently; however, because subgroup error terms are not independent it could be possible that one or two factors would dominate those identified as failing to reach the AYP thresholds.

Evaluative Focus

This exploratory study's focus is guided by the question: Are the decision-rules outlined in Nebraska's AYP misclassifying most schools and district? In decomposing the aforementioned question, two sequential questions were formulated:

(a) What patterns existed in the distribution of AYP classification under the current design?
(b) Do these patterns significantly change using different decision-making rules?

Sample

The initial sample consisted of 140 schools from two large districts. Fifty-nine schools (42.1%) in the sample were from the Lincoln Public School District, while eighty-one schools (57.9%) were extracted from the Omaha Public School District. Ten schools (7.1%) did not have the thirty students required to be evaluated under AYP, thus met the standard by default. Fourteen schools (10.0%) were either special purpose schools or reverted student performance data back to the "home" school, alternative schools, or did not report any data. The final sample utilized one-hundred and twenty-six schools (90%) of the initial sample, although only one-hundred and sixteen schools had those data necessary for AYP evaluations. Table 1 illustrates the school distribution across selected grade configurations from each targeted district.

TABLE 1
School distribution across selected grade configurations

District	Grade Configuration	Schools Included	Schools < 30 Students	Schools Excluded	School Total
Lincoln n = 59					
	High	5	0	7	12
	Middle/Junior High	10	0	0	10
	Elementary	34	2	1	37
Omaha n = 81					
	High	7	0	3	10
	Middle/Junior High	9	0	1	10
	Elementary	51	8	2	61
Total n = 140					
	High	12	0	10	22
	Middle/Junior High	19	0	1	20
	Elementary	85	10	3	98

The final sample represented approximately ten percent of the 2002-2003 population of schools (N = 1,220) in Nebraska. One unique aspect of the population was the large number of small districts and schools located throughout the state. This phenomenon combined with student performance data limited to grades 4, 8, and 11 significantly reduced the number of schools and districts evaluated for AYP in 2002-2003. For example, six hundred and seventy-four (55.2%) of all schools were not evaluated under AYP. Of the five hundred and forty-six schools evaluated, only one hundred and seventy-seven (32.4%) met all AYP conditions, (reading, math, participation, other academic indicators). Of the five hundred and seventeen districts in the state, only one hundred and fifty-nine (30.8%) were evaluated under AYP resulting in only sixty-seven (42.1%) meeting the AYP standards. Tables 2 and 3 compare the AYP evaluated distributions for the sample and population.

TABLE 2
Population vs. sample AYP evaluated school distribution (count)

NCLB Subgroup	**Met AYP state**	**Met AYP sample**	**Not Met AYP state**	**Not Met AYP sample**	**Not Evaluated state**	**Not Evaluated sample**
All Students	**360**	81	**186**	35	**674**	10
American Indian / Alaska Native	**2**	0	**0**	0	**1218**	126
Asian or Pacific Islander	**1**	0	**1**	1	**1218**	125
White, not Hispanic	**325**	71	**131**	10	**764**	45
Black, not Hispanic	**5**	3	**24**	22	**1191**	101
Hispanic	**8**	4	**21**	9	**1191**	113
Economically disadvantaged	**37**	20	**95**	45	**1088**	61
Students with disabilities	**0**	0	**20**	10	**1200**	116
Limited English proficient	**1**	1	**6**	3	**1213**	122

TABLE 3
Population vs. sample AYP evaluated school distribution (percent)

NCLB Subgroup	**Met AYP population**	**Met AYP sample**	**Not Met AYP population**	**Not Met AYP sample**	**Met AYP Difference population vs. sample**
All Students	**65.9**	69.8	**34.1**	30.2	**(3.9)**
American Indian / Alaska Native	**100.0**	0.0	**0.0**	0.0	**100.0**
Asian or Pacific Islander	**50.0**	0.0	**50.0**	100.0	**50.0**
White, not Hispanic	**71.3**	87.7	**28.7**	12.3	**(16.4)**
Black, not Hispanic	**17.2**	12.0	**82.8**	88.0	**5.2**
Hispanic	**27.6**	30.8	**72.4**	69.2	**(3.2)**
Economically disadvantaged	**28.0**	30.8	**72.0**	69.2	**(2.8)**
Students with disabilities	**0.0**	0.0	**100.0**	100.0	**0.0**
Limited English proficient	**14.3**	25.0	**85.7**	75.0	**10.7**

A review of the evaluated AYP distribution suggests the sample and population distributions have a high degree of commonality. Three subgroups have skewed distributions resulting from extremely low n-counts for both the evaluated sample and population. One subgroup, “White, not Hispanic” was statistically higher than the population, suggesting the performance of this subgroup was atypical. AYP status inferences made from the aforementioned subgroup should be considered in terms of how closely the sample represents the population.

Participation Rates

AYP requires students enrolled in school for a full academic year (FAY) to be included in determining AYP status; however, all students are expected to participate in the state's assessment program. Participation rates are computed across nine subgroups for two targeted academic areas, thus requiring a maximum of eighteen cells' status being evaluated for AYP. An eligible cell (EC) is one amenable to AYP evaluation because it contains at least thirty students or at least forty students for cells exclusively representing students with disabilities. Once the number of EC has been determined, the AYP participation threshold of 95% is examined for each EC. Any EC with a participation rate below the established threshold is labeled as a failed cell (FC), which results in the school (or district) AYP status labeled as "NOT MET".

The first exploratory step examined the frequency distribution of eligible cells (EC) across subgroups for reading and mathematics. Each school had a possible distribution range between eighteen and zero eligible cells (ten schools had zero EC). Three grade configurations were used to place schools into categories that reflected the grade-level assessment data. A review of the within category distribution shows that the ratio of elementary schools to other configurations is approximately five to one, suggesting a high degree of centralization occurring from the middle school and beyond. Table 4 provides the EC distribution (school-level) across three grade configurations.

TABLE 4
School distribution across eligible AYP cell categories (participation)

School Level	**0 EC**	**1 EC**	**2 EC**	**4 EC**	**6 EC**	**8 EC**	**9 EC**	**10 EC**	**12 EC**	**14 EC**
High School (n =12)	---	---	---	1	---	6	1	2	1	1
Middle School (n = 19)	---	---	---	2	7	6	---	2	1	1
Elementary School (n = 95)	10	1	12	43	24	5	---	---	---	---
All Schools (n = 126)	*10*	*1*	*12*	*46*	*31*	*17*	*1*	*4*	*2*	*2*

The second exploratory step determined which eligible cells met the AYP standard, then explored the relationship between EC and FC. Reliability studies (Hill, 2001; Hill & DePascale, 2003) have suggested larger, more diverse schools are affected disproportionately by conjunctive decision-making rules. In other words, as the number of eligible cells increase within a school so does the chance of not meeting the AYP standard. The data in Table 5 provides the number of schools within each failed cell (FC) category across three grade configurations.

TABLE 5
School distribution across failed AYP cell categories (participation)

School Level	**0 FC**	**1 FC**	**2 FC**	**3 FC**	**4 FC**	**5 FC**	**6 FC**	**7 FC**
High School (n = 12)	1	---	---	3	2	3	2	1
Middle School (n = 19)	3	2	4	4	---	4	1	1
Elementary School (n = 95)	91	2	2	---	---	---	---	---
All Schools (n = 126)	*95*	*4*	*6*	*7*	*2*	*7*	*3*	*2*

After examining the data from Table 4 and 5, a total of 620 EC existed and of those, one hundred and twelve (18.1%) were categorized as failing cells. The one hundred and twelve FC represented thirty-one schools (24.6%) and these schools were for the most part configured as high and middle schools. Of the twelve high schools, eleven (91.7%) failed the participation standard, followed closely by the sixteen (84.2%) middle/junior high schools. Conversely, only four (4.2%) elementary schools had failing AYP cells. A Pearson product correlation coefficient for the entire sample was used to examine statistically the association between eligible and failed cells. The resulting coefficient ($r = +.6993$, $n = 126$, $p<.01$, two tail) suggested the magnitude FC distribution paralleled that of the EC.

The final exploratory step examined how FC distribution was affected by other states' decision-making rules. Using information from *A Framework for Examining Validity in State Accountability Systems* (Forte-Fast & Hebbler, 2004), various combinations of minimum n-counts, the Texas 50-10%-200 rule, and two alpha levels (ά =.05 and .01) were modeled. Actual participation rates were not reported for the sample; however, a set of synthetic data was produced for those schools with failing cells. These synthetic values ranged from liberal (94%) to conservative (88%) based upon the participation rates reported on the state's writing assessment. Subgroup n-counts were not reported for each school, thus requiring an interpolated dataset from actual enrollment status data. This synthetic dataset was developed to proportionally represent the school-level distribution at the targeted grade and then adjusted to match the reported AYP criterion. The full academic year (FAY) criterion was not imposed on the dataset which under standard AYP decision-rules would have excluded the results of any student not in school all year. The inclusion of FAY data could influence the AYP results in a more conservative direction by an estimated twenty percent. In other words, by including FAY students in the dataset, the results provide a "worst case" scenario as transient student data would increase EC and FC status. A qualitative review of each school's data suggests the subgroup n-counts fit the parameters of known data. Table 6 provides a comparative review of the population and sample's subgroup distribution at the student-level, while Table 7 illustrates the change in AYP participation status across several different decision-making rules.

TABLE 6
Student- level distribution across population vs. sample subgroups (count)

NCLB Subgroup	Population N	Sample n	Population %	Sample %	Percentage Pt. Difference
All students	**283908**	77853	**---**	---	**---**
American Indian/Alaska Native	**4429**	1076	**1.6**	1.4	**0.2**
Asian or Pacific Islander	**4607**	1984	**1.6**	2.5	**(0.9)**
White, Not Hispanic	**228923**	49090	**80.6**	63.1	**17.5**
Black, Not Hispanic	**19843**	16580	**7.0**	21.4	**(14.1)**
Hispanic	**26106**	9123	**9.2**	11.7	**(2.5)**
Economically disadvantaged	**92015**	33826	**32.4**	43.4	**(11.0)**
Students with disabilities	**41244**	11288	**14.8**	14.5	**0.3**
Limited English proficient	**13813**	6883	**5.0**	8.8	**(3.8)**

TABLE 7
School-level distribution of AYP participation rate status across varied decision-making rules

AYP Participation Rate Status	≥ 30; ≥ 45 SWD	≥ 40; ≥ 45 SWD	≥ 50; ≥ 50 SWD	≥ 40; subgroups 50-10%-200	Alpha .05	Alpha .01
Failed (count)	31	28	27	27	12	1
Passed (count)	95	98	99	99	114	125
Failed (percent)	24.6	22.2	21.4	21.4	9.5	0.8
Passed (percent)	75.4	77.8	78.6	78.6	90.5	99.2

The results in Table 7 were based upon a fixed ninety-two percent participation rate being applied to each FC in the dataset. Although the number of schools with FC will fluctuate across the spectrum of possible participation rate distributions, the application of a fixed rate allowed the influence of the decision-making rules to be ascertained. The data suggests that increasing the minimum n-count has very limited influence in reducing the number of schools with failed cells. This phenomenon can be partially explained by the strong relationship between eligible cells and failed cells. In other words, diverse schools with large student populations (middle and high schools) remained eligible for AYP evaluation, while small schools (elementary) had already limited the number of EC. Only five elementary schools had EC associated with more than a two subgroups beyond the "whole school" subgroup. However, applying a confidence interval reduces significantly the number of FC and consequently the number of schools failing the AYP participation standard. The data in Table 7 demonstrates clearly the impact of applying confidence intervals, especially at the .01 alpha level.

Performance Rates

AYP requires the academic performance (achievement) of students on statewide assessments be used to evaluate schools and districts, in the aggregate, and across targeted subgroups. A subgroup (cell) is amenable to AYP evaluation when the minimum number of student threshold, similar to the method used for participation rate, has been attained for both reading and mathematics. For eligible cells (EC), a targeted percent of students must be proficient in reading and mathematics across all eligible cells. EC with proficiency levels below the established thresholds are determined to be failed cells (FC) and labeled as "NOT MET" AYP. In Nebraska, AYP proficiency thresholds were established across two academic areas and three grade configurations. The other academic indicator (OAI) used for elementary and middle/junior high schools is the state administered writing test. High schools are required under NCLB to use graduation rates as their OAI. Table 8 provides an overview of the different performance thresholds used for AYP decisions.

TABLE 8
Performance thresholds across performance indicators (percent)

School Level	Reading Threshold	Math Threshold	Writing Threshold	Graduation Threshold
High School (Grade 11)	66	62	---	83.97
Middle School (Grade 8)	61	58	62	---
Elementary School (Grade 4)	62	65	61	---

The first exploratory step focused the EC distribution across subgroups for reading and mathematics using the same method outlined for participation rates. It was hypothesized that the participation and performance distributions for eligible cells would be identical, which was confirmed by comparing Tables 4 and 9.

TABLE 9
School distribution across eligible AYP cell categories (performance)

School Level	0 EC	1 EC	2 EC	4 EC	6 EC	8 EC	9 EC	10 EC	12 EC	14 EC
High School (n =12)	---	---	---	1	---	6	1	2	1	1
Middle School (n = 19)	---	---	---	2	7	6	---	2	1	1
Elementary School (n = 95)	10	1	12	43	24	5	---	---	---	---
All Schools (n = 126)	*10*	*1*	*12*	*46*	*31*	*17*	*1*	*4*	*2*	*2*

The second exploratory step investigated the relationship between AYP cell eligibility and performance status. As seen with the participation rate indicator, the AYP performance status was expected to be associated strongly with the number of eligible cells within a school. A total of six hundred and twenty (27.3%) out of 2,268 possible cells were eligible for AYP performance evaluation. Of the six hundred and twenty EC, two hundred and ten (33.9%) cells failed to reach the targeted threshold, thus resulting in the cell being deemed as a failed cell. The FC group comprised of fifty-four schools, mostly high and middle schools. Of the high schools, twelve out of twelve (100%) had at least one FC. For middle schools, fourteen out of nineteen (73.7%) schools failed to reach the AYP standard. A relatively large number of elementary schools, twenty-eight out of ninety-five (29.5%) had failing cells, a significant increase over the participation rate indicator. As expected, the number of eligible cells continued to demonstrate a high correlation to the number of failed cells (r = +.6776, n = 126, *p*<.01, two tail).

TABLE 10
School distribution across failed AYP cell categories (performance)

School Level	**0 FC**	**1 FC**	**2 FC**	**3 FC**	**4 FC**	**5 FC**	**6 FC**	**7 FC**	**8 FC**	**9 FC**	**10 FC**	**11 FC**
High School (n = 12)	---	---	1	---	1	1	4	1	1	1	1	1
Middle School (n = 19)	5	2	5	4	---	---	2	1	---	---	---	---
Elementary School (n = 95)	67	7	9	1	4	1	5	---	1	---	---	---
All Schools (n = 126)	*72*	*9*	*15*	*5*	*5*	*2*	*11*	*2*	*2*	*1*	*1*	*1*

The other academic indicators (OAI) were also evaluated during this stage of the study. For high schools, the OAI was an estimated four-year, high school completion rate with an established threshold set at 83.97%. The NCLB method estimates the completion rate by adjusting for dropouts at each grade-level (9-12); however, it should not be misinterpreted as a longitudinal tracking of a ninth grade cohort across four years. Estimated completion rates should be considered as approximations of the true cohort completion rate, especially due to exogenous factors (*i.e.,* student mobility) which can influence error associated with conducting four, independent samples. The OAI data suggests that twelve out of twelve (100%) of all high schools had failing AYP cells. The OAI data for elementary and middle schools was extracted from the statewide writing assessments with established proficiency thresholds of sixty-two percent (62%) and sixty-one percent (61%) respectively. The results suggested that of the nineteen middle schools, seven (36.8%) schools had at least one FC. Forty out of ninety-five (42.1%) elementary schools had failed cells, which was an increase of twelve schools. In summary, the OAI had a greater influence on FC status than did the reading and mathematics performance indicators.

The final exploratory step applied a confidence interval using the method outlined by states (Erpenbach, Forte-Fast & Potts, 2003). Because limited performance data were reported at the subgroup level, an interpolated dataset was constructed. This synthetic dataset was designed to fit known parameters across subgroups for each academic indicator to include the presence of performance gaps between subgroups. For example, the percent of FC across the "White, not Hispanic" subgroup was only eight percent in comparison to the 100% FC distribution found among the students with disabilities (SWD) cells. After computing the odds ratio between targeted subgroups, the odds of an eligible white subgroup cell failing the AYP standard was determined to be thirteen times lower than that of a "Hispanic" cell and thirty-two times lower than a "Black, not Hispanic" cell. Further, the proportion of failed Hispanic cells was approximately that of economically disadvantaged cells. These known distributions were used to replicate vertically (within an academic indicator) and horizontally (between academic indicators). The resulting synthetic dataset was qualitatively reviewed to ensure the values did not exceed those known parameters.

Using the synthetic dataset, confidence intervals (CI) were computed using alpha levels of .05 and .01. The upper limit of each interval was then added to the proportion of proficient students and reevaluated against the applicable threshold. The interval data were constructed by estimating the standard error of the proportion based upon a normal distribution. The resulting distribution of failed cells was reduced to forty-five (35.7%) after applying a 95% confidence interval. The elementary school distribution was most influenced in that eight fewer schools were identified as not having FC. Middle schools reduced the number of schools with FC by three yet high schools had no change. The aforementioned procedure was repeated using a 99% confidence interval; however, no change was observed in the high and middle school FC distribution. The shift in distribution appeared exclusive to elementary schools by reducing the number of FC schools from twenty-two to fifteen. As expected, the 95% confidence interval reduced the number of middle schools failing the OAI by one and the 99% interval did not influence the distribution. The elementary school distribution revealed a significant shift downward, forty to fifteen schools with FC, when the 95% CI was applied. Four elementary schools were eliminated using the higher 99% confidence interval.

Tables 11 and 12 provide the distributions across failed AYP cell categories using the upper limits of the 95% and 99% confidence intervals. Tables 13 and 14 provide an overview of the number of schools identified using different alpha levels. An alternative CI method, based upon the binomial distribution and less influenced by extreme proportions and n-counts (Coladarci, 2003, Ghosh, 1979), was applied to the sample data. This method resulted in no significant change in the sample's distribution of failed cells. A post-interval (95%), academic area correlation coefficient was calculated between the number of failed reading and mathematics cells. The results suggests a significantly positive relationship (r = +.5720, n = 126, *p*>.01, two-tail) between the two academic areas. This relationship strengthened (r = +.7033, n = 126, *p*>.01, two-tail) when the 99% confidence interval was applied to the failed cell values, suggesting an increased association.

TABLE 11
School distribution across failed AYP cell categories (95% confidence interval)

School Level	**0 FC**	**1 FC**	**2 FC**	**3 FC**	**4 FC**	**5 FC**	**6 FC**	**7 FC**	**8 FC**	**9 FC**	**10 FC**
High School (n = 12)	---	---	2	1	1	1	3	---	1	2	1
Middle School (n = 19)	8	1	6	1	---	1	1	1	---	---	---
Elementary School (n = 95)	73	8	5	4	3	1	1	---	---	---	---
All Schools (n = 126)	*81*	*9*	*13*	*6*	*4*	*3*	*5*	*1*	*1*	*2*	*1*

TABLE 12
School distribution across failed AYP cell categories (99% confidence interval)

School Level	**0 FC**	**1 FC**	**2 FC**	**3 FC**	**4 FC**	**5 FC**	**6 FC**	**7 FC**	**8 FC**	**9 FC**	**10 FC**
High School (n = 12)	---	1	3	---	1	2	1	---	1	2	1
Middle School (n = 19)	8	3	4	1	---	2	---	1	---	---	---
Elementary School (n = 95)	80	8	3	2	1	---	1	---	---	---	---
All Schools (n = 126)	*88*	*12*	*10*	*3*	*2*	*4*	*2*	*1*	*1*	*2*	*1*

TABLE 13
School failing AYP across differing confidence intervals (reading and mathematics)

School Level	No CI	95% CI	99% CI	Overall Change
High School (n = 12)	12	12	12	0
Middle School (n = 19)	14	11	11	3
Elementary School (n = 95)	28	22	15	13
All Schools (n = 126)	*54*	*45*	*38*	*16*

TABLE 14
School failing AYP across differing confidence intervals (Other Academic Indicator)

School Level	No CI	95% CI	99% CI	Overall Change
High School (n = 12)*	12	12	12	0
Middle School (n = 19)	7	6	6	1
Elementary School (n = 95)	40	15	11	29
All Schools (n = 126)	*59*	*33*	*29*	*30*

* Confidence intervals were estimated using the weighted sample mean

Summary

The evaluative focus of this study reflected upon the current distribution of eligible and failed AYP cells along with how the distribution is influenced through the use of alternative, decision-making rules. A key finding was the strong association between the number of eligible and failed cells. This positive relationship was unaffected by changes in the minimum n-counts for either the participation or performance indicators. The application of confidence intervals reduced the number of with cells failing to attain the participation rate threshold, which may be a result of limited variability. In other words, most schools were closer to the participation threshold than to the performance thresholds. Tables 13 and 14 revealed changes to the elementary, FC distribution at rates much higher than those observed in the other grade configurations. The writing thresholds (OAI) resulted in a high number of elementary schools gaining a FC; however, the application of confidence intervals considerably reduced the overall number of schools. High schools showed no distribution changes that would have eliminated all FC from any school's overall distribution. Unlike the distribution of participation rates, performance rates across subgroup and content areas have extreme cell values. These cell values are significantly below the established thresholds and when combined with large n-counts produce confidence intervals unable to attain the applicable thresholds. For example, the reading performance of SWD in high and middle schools was approximately half the distance to the targeted threshold. In other words, the odds of having an eligible SWD cell in a school not being labeled as a FC were zero. This finding, along with others in this exploratory study, provides important information about Nebraska's accountability system.

REFERENCES

Coladarci, T. (2003). *Gallup goes to school: The importance of confidence intervals for evaluating "adequate yearly progress" in small schools.* A Policy Brief of the Rural School and Community Trust, Washington, DC: Rural School and Community Trust.

Erpenbach, W. J. (2003). *Statewide educational accountability under NCLB*. Washington, DC: Council of Chief State School Officers.

Forte-Faste, E., & Hebbler, S. (2004). *A framework for examining validity in state accountability systems* Washington, DC: Council of Chief State School Officers.

Ghosh, B. K. (1979). A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association, 74*, 894-900.

Hill, R. (2001). Issues *related to the reliability of school accountability scores.* Report on the reliability lecture from the 2000 Annual Edward F. Reidy Interactive Lecture Series. Dover, NH: National Center for the Improvement of Educational Assessment.

Hill, R., & DePascale, C. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practices, 22*(3), 12-20.

Marion, S.F., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decision in determining adequate yearly progress.* Washington, DC: Council of Chief State School Officers.